**Procedure No. 5-06C**

# AUDIT SAMPLING GUIDELINE

## *DEFINITION OF AUDIT SAMPLING TERMS*:

**Attribute**: The qualitative characteristic of a population element. For example, the classification of vouchers as to whether or not they were properly signed. Sampling for attributes uses two-fold or binomial classifications: "yes" or "no."

**Audit Sampling**: The application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class.

**Confidence Interval**: A range of values for a sample statistic wherein the actual population value is believed to lie. The interval is computed on the basis of a known sample value, a desired precision and a specified level of confidence. For example, assume that the results of a 95% confidence level sample produce a confidence interval whereby the lower confidence limit is $150 and the upper confidence limit is $325. This means that if the auditor were to repeatedly draw random samples of a certain size (e.g., several samples of 80 items) and calculate the confidence interval at the 95% level, then about 95% of the intervals would be expected to encompass the actual population value.

**Confidence Level or Degree of Assurance**: The confidence level indicates the degree of assurance (probability) that the results of a sample are reasonable estimates of specific population characteristics. Confidence levels are usually expressed in percentages such as 90 or 95 per cent. A 95% confidence level means that if repeated samples were selected, the actual value would fall within the confidence intervals about 95% of the time.

**Frequency Distribution**: The classification of numerical data according to size or magnitude (how many or how much). A population whose elements are classified according to some quantitative characteristic (e.g., dollar values of purchase order, of invoices, etc) may be described by a frequency distribution. A symmetrical distribution is a frequency distribution that can be portrayed by a normal (bell) curve. A skewed distribution is a frequency distribution which extends further in one direction than in another. An example of a skewed distribution would be one that involved 500 invoices totaling $1,000,000 with 5 of the invoices amounting to $100,000 each. By examining all the high-value invoices, and thereby removing them from the skewed distribution, the remaining invoices, assuming no further extremes, would reflect a normal distribution lending itself to random sampling.

**Mean**: The sum of the population values divided by the number of items in the population. Thus, if the population consists of 7 accounts with balances of $3, $6, $7, $10, $13, $14 and $17, the mean would be $10 ($70/7). The mean is also called the arithmetic mean and the average.

**Median**: A central value which divides an array of a set of data (numerically ordered by magnitude) so that one half of the items are the same as or larger than it, and one half of the items are the same as or smaller than it.

**Mode**: The value which occurs most frequently in a set of data. For a set in which each value occurs only once, there is no mode. In another set of data, there may be more than one mode when two values occur the same number of times and are the most frequent.

**Nonsampling risk**: Results from uncertainties that are not due to sampling. For example: (1) Incorrect audit procedures for a given objective or (2) Nonrecognition of errors.

**Population**: Also known as **Universe** or **Field**. The aggregate or entirety of the items or units about which information is desired. The population excludes individually significant items that the auditor has decided to test 100 percent or other items that will be tested separately. For sample results to be reliable, the sample must be selected from the complete population.

**Precision**: The range within which the estimate of the population value or characteristics will fall at the confidence level. It is a range or tolerance and is usually expressed as a plus-or minus percentage, such as +/- 3 percent, or as an amount, such as +/- $1,000. For example, an auditor may conclude that there is a 90 percent probability, or confidence, that the average value of an account is within $50, either way of the sample average of $700. In turn, there is a 10 percent risk that the average value is greater than $750, or less than $650.

**Probability**: The ratio of the frequency of certain events to the frequency of all the possible events in a series or set. In other words, the number of times that something can occur in a specific way, as compared with the number of times it can happen in all possible ways. This is usually expressed as a decimal ratio which can be converted to a percentage by multiplying by 100.

**Sampling Risk**: Results from the possibility that if a test is restricted to a sample, the conclusions reached may be different than the conclusions that may result if the entire population is examined.

**Standard Deviation**: Measures the degree to which individual values in a list vary from the mean (average) of all values in the list. The lower the standard deviation, the less individual values vary from the mean, and the more reliable the mean. A standard deviation of 0 indicates that all values in the list are equal. To obtain the standard deviation, one must calculate the difference between the value of each individual item in the population and the population mean, square these differences, add them, divide the sum by the total number of items, and finally extract the square root.

**Variability**: A measure designed to describe the scatter or dispersion of a Frequency Distribution.

**Variable**: A quantitative characteristic of a elements (items) of a population which may vary from the observation of one item to another.

**Variance**: The square of the standard deviation.

## STATISTICAL SAMPLING

Statistical sampling is based on the assumption that, within a given confidence (assurance) level and allowance for sampling risk, a randomly selected sample of items from a population will reflect the same characteristics that occur in the population. Therefore, auditors may draw valid conclusions based on data derived from a relatively small sample of the total population. Statistical sampling allows the auditor to provide a mathematical measurement of the degree of uncertainty that results from examining only part of the population.

Following is a description of attributes and variables sampling plans with recommended use of those plans. The auditor may apply both plans to the same sample selected.

### Attributes Sampling

Recommended Use: To estimate the qualitative characteristics of a population obtaining "yes or no" answers with a measurable degree of reliability.

In attributes sampling, the auditor tests a sample for qualitative characteristics. The answer is either "yes" or" no". The following general steps are followed in attributes sampling in tests for compliance with prescribed internal control procedures:

1. Determine the precision.
2. Determine the confidence level.
3. Determine the expected population deviation rate.
4. Compute the sample size using sample size tables (available in statistics textbooks), computer software, or the sample size formula (available in statistics textbooks) for attributes sampling.
5. Document the sampling procedure.
6. Select and audit the sample items in terms of whether they exhibit ("yes" or "no") the qualitative characteristics being tested.
7. Evaluate the sample results.
8. Reach an overall conclusion.
9. Document the sample results and conclusions reached.

### Variables Sampling

Recommended Use: To estimate the quantitative characteristic of a population (i.e., dollars, time spans, weights, or other quantitative characteristics) with a measurable degree of reliability.

In variables sampling, the auditor tests a sample to estimate the quantitative characteristics of a population. The following general steps are followed in variables sampling.

1. Determine the objectives of the test.
2. Define the population.
3. Determine the confidence level.
4. Determine the precision.

5. Determine the expected standard deviation of the population.

6. Compute the sample size using sample size tables (available in statistics textbooks), computer software, or the sample size formula (available in statistics textbooks) for variables sampling.

7. Document the sampling procedure.

8. Select and audit the sample items in terms of obtaining an estimate of a quantitative value that is representative of the population based on the value of the sample items.

9. Evaluate the sample results.

10. Reach an overall conclusion.

## STATISTICAL SAMPLE SELECTION METHODS

The following is a list of five statistical sample selection methods available to the auditor with the recommended use for each of the sample selection methods:

**Stratified Sampling**:

Recommended Use: When the population is composed of items which vary significantly in size, either amount or characteristic.

An auditor should attempt to identify wide variations in size, either in amount or characteristic of items making up a population. If wide variations are present, the auditor should consider stratified sampling. Using this method, the auditor first segregates the population into two or more distinct, sub-populations called stratum. The largest, or most expensive, or most significant items in a population can either be 100% examined if the total sub-population is defined small enough, or it can be examined by a separate random sample. A separate random sample is made of each of the remaining stratum. Unless the sub-population is to be sampled 100%, the sample size from each stratum should be at least 30, in order to avoid problems with precision calculation. The results of the several samplings from each of the different stratum may then be combined into an overall estimate for the entire population.

Reasons for using a stratified sample include:

1. To gain sampling efficiency, i.e., to obtain a lower sampling error with the same sample size as compared with unrestricted random sampling.

2. To give special attention to certain categories within the population.

3. To offset the effect of extreme values (skewed distributions.

**Random Number Sampling**:

Recommended Use: Where each of the items in the population is or can be easily numbered.

The auditor selects a random sample by matching random numbers generated by a computer or selected from a random number table with, for example, the document number. With this method, every item in the population has the same probability of being selected as every other item in the population. In a stratified population, where some of the stratum are sampled with greater intensity than others, each item in the entire population does not necessarily have an equal chance of selection.

**Interval Sampling**:

Recommended Use: Where items are not or cannot be easily numbered.

A method by which items are selected from the population in such a way that there is a uniform interval between sample items. The first item in the series must be chose at random and then every "n"th item is chosen to result in the desired sample size.

**Cluster Sampling**:

Recommended Use: When a population is so dispersed that interval sampling or unrestricted random number sampling would be burdensome.

Cluster sampling is the method of sampling whereby the population is formed into groups or "clusters" of items. The first step is to make a random selection of clusters to include in the sample. Then the items within the selected clusters may be randomly selected and sampled. This is called Multi-stage Sampling. For example, in surveying a large number of tool cribs, audit expense may be reduced by the use of cluster sampling: First, a random selection is made of the cribs, and then a random selection is made of the tool records within the cribs that were randomly selected. Cluster Sampling is commonly used to get the most precise results from a fixed budget, however it is not as precise as Random Sampling.

**Computer Sampling**:

Recommended Use: When the population is, or easily can be, recorded on disk, magnetic tape or other machine media.

Computer-assisted sampling can be divided into two general areas:

1. Using either spreadsheet or audit software with either hardcopy or computer media (e.g., disk)

2. Using audit software, such as ACL (available in the City Auditor’s Office).

## NONSTATISTICAL SAMPLE SELECTION METHOD

Nonstatistical sampling does not allow the auditor to extrapolate findings to the entire population. It does allow the auditor, to make positive or negative statements about specific items in the population. (NOTE: SAS No. 39 allows the auditor to use nonstatistical sampling in performing tests of controls.)

The following is the nonstatistical sample selection method available to the auditor with recommended use for the method.

**Judgment Sampling:**

Recommended Use: To use samples for the purpose of obtaining information that need not be attributed to the entire population with measured reliability.

The selection of those items which in the judgement of the auditor best suits the audit objective. The degree to which the judgment sample represents the population cannot be demonstrated mathematically. It therefore, should not be used as representative of the population.

Judgment sampling may be used to select examples of deficiencies to support the auditor's contention that the system is weak. It may also be use to make a directed search for defective items to confirm or support the auditor's belief that the system is not properly identifying defective items.

Judgment sampling may also be used to document that a system of internal controls is functioning properly and that a statistical sample is not necessary.

## IF POSSIBLE, TEST THE ENTIRE POPULATION

For certain types of audit tests, the audit software ACL, which is available in the City Auditor’s Office, allows the auditor to review the entire population. For example, the auditor can test the entire database for the following:

- Duplicate records
- Incorrectly calculated values
- Dollar values or rates other than authorized
- Invalid codes
- Blank fields
- Unreadable information